DOCUMENT RESUME

ED 336 427 TM 017 269

AUTHOR Brennan, Robert L.
TITLE Congeneric Models and Levine's Linear Equating Procedures.
INSTITUTION American Coll. Testing Program, Iowa City, Iowa.
REPORT NO ACT-RR-90-12
PUB DATE Dec 90
NOTE 33p.; Revised version of a paper presented at the Annual Meeting of the Psychometric Society (Princeton, NJ, June 2-7, 1990).
AVAILABLE FROM ACT Research Report Series, P.O. Box 168, Iowa City, IA 52243.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equated Scores; *Equations (Mathematics); *Mathematical Models; Research Design; Testing Problems; Test Length; *True Scores
IDENTIFIERS Anchor Tests; Congeneric Tests; *Levine Equating Method; *Linear Equating Method

ABSTRACT

In 1955, R. Levine introduced two linear equating procedures for the common-item non-equivalent populations design. His procedures make the same assumptions about true scores; they differ in terms of the nature of the equating function used. In this paper, two parameterizations of a classical congeneric model are introduced to model the variables in the Levine procedures for the external and internal anchor cases. The models differ in the constraints imposed on certain effective test length parameters, as well as assumptions made about one covariance term. This modelling leads to simple expressions for true-score variances, reliabilities, and Angoff error variances. Applying these two parameterizations of the classical congeneric model with the Levine assumptions leads to general equations for both Levine procedures and both external and internal anchor cases that involve ratios of the effective test length parameters. The role of synthetic population weights for both Levine procedures is considered, along with an alternative interpretation of one Levine procedure. Multiple equations, one table, and one appendix support the discussion. A 19-item list of references is included. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ACT Research Report Series 90-12

ED 336427

Congeneric Models and Levine's Linear Equating Procedures

Robert L. Brennan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. FERGUSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

December 1990

BEST COPY AVAILABLE

ACT

TM017269

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

© 1990 by The American College Testing Program. All rights reserved.

Congeneric Models and Levine's Linear Equating Procedures

Robert L. Brennan

Table of Contents

Page

Abstract.....ii

Terminology.....2

Some Results for the Classical Congeneric Model.....3
- Tests X and V Disjoint.....4
- Test V Included in Test X.....6
- Comments.....8

Levine Observed-Score Method.....8
- Assumptions.....9
- General Results.....10
- External Anchor.....11
- Internal Anchor.....12
- Comment.....12

Levine True-Score Method.....13
- General Results.....14
- External Anchor.....16
- Internal Anchor.....16
- First-Order Equity.....17

Summary and Discussion.....20

References.....24

Appendix.....26

Author Note.....27

Table 1.....28

Abstract

In 1955 Levine introduced two linear equating procedures for the common-item nonequivalent-populations design. His two procedures make the same assumptions about true scores; they differ in terms of the nature of the equating function employed.

In this paper two parameterizations of a classical congeneric model are introduced to model the variables in the Levine procedures for the external and internal anchor cases. The models differ in the constraints imposed on certain effective test length parameters, as well as assumptions made about one covariance term. This modeling leads to simple expressions for true-score variances, reliabilities, and the so-called "Angoff error variances."

Applying these two parameterizations of the classical congeneric model with the Levine assumptions leads to general equations (for both of the Levine procedures and both the external and internal anchor cases) that involve ratios of the effective test length parameters. This presentation facilitates interpretation.

The role of synthetic population weights for both Levine procedures is considered, along with an alternative interpretation of one of Levine's procedures.

Congeneric Models and Levine's Linear Equating Procedures

Levine (1955) introduced two linear equating methods for a design in which two non-equivalent populations take different forms of a test with a common set of equating items, or anchor. Levine referred to his two methods as major-axis procedures. Angoff discussed these methods in his 1971 chapter on Scales, norms, and equivalent scores in the second edition of Educational Measurement. Angoff's chapter was reprinted in 1984 by the Educational Testing Service. Other authors who have treated one or both of these methods include Woodruff (1986, 1989), Kolen and Brennan (1987), Petersen, Kolen, and Hoover (1989), MacCann (1990), and Hanson (1990).

Levine's methods make assumptions about true scores and error scores. Consequently, to apply these methods, it is necessary to model the relationships among observed, true, and error scores. In this paper, a particular version of a congeneric model is employed in which the error variances are assumed to follow classical assumptions. Actually, two parameterizations of the model are employed--one that is associated with the common items constituting an external anchor, and the other for an internal anchor in which the common items are part of the full length forms.

For both of Levine's methods, this modeling leads to general equations that involve ratios of certain effective test length parameters. These parameters aid in presenting and interpreting results. It is also shown that Angoff's (1984, pp. 114-115) results for Levine's methods can be obtained from the results presented here.

The paper ends with a discussion of an alternative conception of one of Levine's methods, followed by a consideration of other issues of interpretation and possible future research.

Terminology

Terminology employed with Levine's procedures has become somewhat confused or, at best, inconsistent in recent years. In particular, Levine originally distinguished between his procedures in terms of their presumed appropriateness for equally reliable and unequally reliable tests. However, Woodruff (1986), Kolen and Brennan (1987), and Hanson (1990) have all noted that Levine's results can be derived without making any assumptions about the reliabilities of the tests involved. Rather, the distinguishing difference in the derivations of the procedures is that the so-called "equally reliable" method is an observed-score equating procedure, whereas the so-called "unequally reliable" method uses observed scores in a linear relationship between true scores for the two forms. Here, therefore, to avoid perpetuating the impression that Levine's procedures make reliability assumptions, the procedures will be referred to as Levine's observed-score and true-score equating procedures. (Admittedly, the phrase "true-score equating" is a bit inaccurate because, as noted above, it is actually observed scores that are used in a true-score relationship, but this inconsistency seems slight compared to the potential misunderstanding inherent in the phrases "equally-reliable" and "unequally-reliable.")

Also, as noted by Woodruff (1989), a distinction can be drawn between the results for Levine's procedures as expressed by Levine (1955), and a particular case of Levine's results that Angoff (1984) provides. In terms of formulas, Levine's "general" results and Angoff's version of Levine's results can be distinguished by the fact that Levine (1955) typically expresses true-score variance as observed-score variance times reliability, without specifying a specific reliability coefficient. By contrast, Angoff's versions of Levine's results are based on specific reliability coefficients that are

derived using "Angoff's error variances" (see Angoff, 1953). In this paper, whenever there is the potential for confusion with respect to this distinction, "Levine-Angoff" will be used to designate Angoff's (1984) results for Levine's methods. Largely, this paper deals with alternative and somewhat more general derivations and presentations of the Levine-Angoff results--a presentation intended to aid interpretation.

Some Results for the Classical Congeneric Model

Let X and V designate observed scores for two tests or sets of items. For the congeneric model, X and V are decomposed as follows:

$$X = T_X + E_X = (\lambda_X T + \delta_X) + E_X \quad \text{and} \tag{1}$$

$$V = T_V + E_V = (\lambda_V T + \delta_V) + E_V \; . \tag{2}$$

A particular version of the congeneric model arises when it is further specified that

$$\sigma^2(E_X) = \lambda_X \sigma^2(E) \quad \text{and} \tag{3}$$

$$\sigma^2(E_V) = \lambda_V \sigma^2(E) \; . \tag{4}$$

This special version will be called here the classical congeneric model. It is discussed by Feldt (1975), Feldt and Brennan (1989, pp. 111-112) and Woodruff (1986), among others. The word "classical" is used here to indicate that the error variances are proportional to the "effective" test lengths, λ_X and λ_V . In this sense, this model is closer to the traditional classical test theory model than would be the case if Equations 3 and 4 did not hold.

Discussed next are two parameterizations of this classical congeneric model. These parameterizations differ with respect to constraints imposed on the λ's and assumptions about $\sigma(E_X, E_V)$. The first case is for tests X and V disjoint and will be applied later to external anchor equating; the second

case is for V included in X and will be applied later to internal anchor equating.

Tests X and V Disjoint

Suppose that tests X and V are disjoint in the sense that they contain no common items. To represent this case, we assume that errors have an expectation of zero, all covariances between true and error scores are zero and, since X and V are distinct,

$$\sigma(E_X, E_V) = 0 . \tag{5}$$

For identifiability purposes we impose the usual constraint

$$\lambda_X + \lambda_V = 1 . \tag{6}$$

(It is also usual to impose the constraint $\delta_X + \delta_V = 0$, but doing so is not required for the following derivations.)

For this model, the variances and covariances are easily determined:

$$\sigma^2(X) = \lambda_X^2\sigma^2(T) + \lambda_X\sigma^2(E) , \tag{7}$$

$$\sigma^2(V) = \lambda_V^2\sigma^2(T) + \lambda_V\sigma^2(E), \quad \text{and} \tag{8}$$

$$\sigma(X,V) = \lambda_X\lambda_V\sigma^2(T) . \tag{9}$$

Further, letting A = X + V (recall that X and V consist of non-overlapping sets of items) it is easy to show that

$$\sigma^2(A) = \sigma^2(T) + \sigma^2(E) .$$

To derive λ_X in terms of variances and covariances, note that

$$\begin{aligned}\sigma^2(X) + \sigma(X,V) &= \lambda_X^2\sigma^2(T) + \lambda_X\sigma^2(E) + \lambda_X\lambda_V\sigma^2(T) \\ &= \lambda_X[(\lambda_X + \lambda_V)\sigma^2(T) + \sigma^2(E)] \\ &= \lambda_X[\sigma^2(T) + \sigma^2(E)] \\ &= \lambda_X\sigma^2(A) .\end{aligned}$$

It follows that the effective test length for X is

$$\lambda_X = [\sigma^2(X) + \sigma(X,V)]/\sigma^2(A) \tag{10}$$

$$= \sigma(X,X+V)/\sigma^2(A)$$

$$= \sigma(X,A)/\sigma^2(A)$$

$$= \alpha(X|A) , \tag{11}$$

where $\alpha(X|A)$ is the slope of the linear regression of X on A. Similarly, the effective test length for V is

$$\lambda_V = [\sigma^2(V) + \sigma(X,V)]/\sigma^2(A) \tag{12}$$

$$= \alpha(V|A) . \tag{13}$$

Note that, $[\sigma^2(X) + \sigma(X,V)]/\sigma^2(A)$ and $[\sigma^2(V) + \sigma(X,V)]/\sigma^2(A)$ are both (relative) effective weights, as defined by Wang and Stanley (1970). Hence, the effective test lengths λ_X and λ_V are also interpretable as (relative) effective weights, as well as slopes of X (or V) on A = X + V.

For this classical congeneric model, using Equation 9 and noting that $\sigma^2(T_V) = \lambda_V^2\sigma^2(T)$, we obtain

$$\sigma^2(T_V) = (\lambda_V/\lambda_X)\ \sigma(X,V) . \tag{14}$$

Consequently, the reliability of test V is

$$\rho(V,V') = \sigma^2(T_V)/\sigma^2(V)$$

$$= (\lambda_V/\lambda_X)\ \sigma(X,V)/\sigma^2(V) . \tag{15}$$

Also, using Equation 14 (and then Equations 10 and 12) the variance of the errors associated with test V is found to be

$$\sigma^2(E_V) = \sigma^2(V) - (\lambda_V/\lambda_X)\ \sigma(X,V) \tag{16}$$

$$= \frac{\sigma^2(X)\sigma^2(V) - [\sigma(X,V)]^2}{\sigma^2(X) + \sigma(X,V)} . \tag{17}$$

For test X, equations for true-score variance, reliability, and error variance can be obtained by interchanging X and V in Equations 14-17 resulting in

$$\sigma^2(T_X) = (\lambda_X/\lambda_V)\ \sigma(X,V)\ , \tag{18}$$

$$\rho(X,X') = (\lambda_X/\lambda_V)\ \sigma(X,V)/\sigma^2(X), \quad \text{and} \tag{19}$$

$$\sigma^2(E_X) = \sigma^2(X) - (\lambda_X/\lambda_V)\sigma(X,V) \tag{20}$$

$$= \frac{\sigma^2(X)\sigma^2(V) - [\sigma(X,V)]^2}{\sigma^2(V) + \sigma(X,V)}\ . \tag{21}$$

Equations 17 and 21 are the usual expressions for the so-called "Angoff error variances" for tests V and X, respectively, for X and V disjoint (see Angoff, 1953; Petersen et al., 1989, p. 254).

Test V Included in Test X

The previous section considered the case of tests X and V containing no common items. In equating terminology this is the case associated with V being an external anchor. Suppose now that test V is an internal anchor--i.e., all of the items in test V are included in test X. In this case, the classical congeneric model Equations 1-4 still apply, and we assume that errors have an expectation of zero and all covariances between true and error scores are zero. However, we replace the constraint in Equation 6 with

$$\lambda_X = 1\ . \tag{22}$$

(For completeness it is typical to specify the constraint $\delta_X = 0$, but doing so is not necessary for deriving the results that follow.) Setting $\lambda_X = 1$ merely specifies that, when V is included in X, the full-length test, X, has an effective length of 1. Consequently, for this model we let $T_X = T$ and $E_X = E$, and Equation 1 can be written

$$X = T_X + E_X = T + E\ .$$

Equation 5, $\sigma(E_X,E_V) = 0$, is not valid for this case, however. Rather, since V is included in X, only the covariance between V and the non-common part of X is zero. Therefore,

$$\sigma(E_X,E_V) = \sigma(E,E_V) = \sigma(E_V,E_V) = \sigma^2(E_V) = \lambda_V\sigma^2(E)\ . \tag{23}$$

For this internal anchor classical congeneric model, the variances and covariances are easily found to be

$$\sigma^2(X) = \sigma^2(T) + \sigma^2(E) , \tag{24}$$

$$\sigma^2(V) = \lambda_V^2\sigma^2(T) + \lambda_V\sigma^2(E), \text{ and} \tag{25}$$

$$\sigma(X,V) = \lambda_V\sigma^2(T) + \lambda_V\sigma^2(E) \tag{26}$$

$$= \lambda_V\sigma^2(X) . \tag{27}$$

From Equation 27, it is clear that

$$\lambda_V = \sigma(X,V)/\sigma^2(X) = \alpha(V|X) . \tag{28}$$

Again we find that the effective test length parameter λ_V is a slope.

Recalling that $\sigma^2(T_V) = \lambda_V^2\sigma^2(T)$, and solving Equations 25 and 26 simultaneously, we obtain

$$\sigma^2(T_V) = \frac{\lambda_V}{1 - \lambda_V}[\sigma(X,V) - \sigma^2(V)] . \tag{29}$$

Consequently, the reliability of test V is:

$$\rho(V,V') = \frac{\lambda_V}{1 - \lambda_V} \quad \frac{\sigma(X,V) - \sigma^2(V)}{\sigma^2(V)} . \tag{30}$$

Also, using Equation 29 (and then Equation 28)

$$\sigma^2(E_V) = \frac{\sigma^2(V) - \lambda_V\sigma(X,V)}{1 - \lambda_V} \tag{31}$$

$$= \frac{\sigma^2(X)\sigma^2(V) - [\sigma(X,V)]^2}{\sigma^2(X) - \sigma(X,V)} . \tag{32}$$

Since $\sigma^2(T_V) = \lambda_V^2\sigma^2(T) = \lambda_V^2\sigma^2(T_X)$, it follows from Equation 29 that

$$\sigma^2(T_X) = \frac{\sigma(X,V) - \sigma^2(V)}{\lambda_V(1 - \lambda_V)} , \tag{33}$$

and the reliability of test X is

$$\rho(X,X') = \frac{\sigma(X,V) - \sigma^2(V)}{\lambda_V(1 - \lambda_V)\sigma^2(X)} . \tag{34}$$

Since $\sigma^2(E_V) = \lambda_V\sigma^2(E) = \lambda_V\sigma^2(E_X)$ and $\lambda_V = \sigma(X,V)/\sigma^2(X)$, it follows from Equations 31 and 32, respectively, that

$$\sigma^2(E_X) = \frac{\sigma^2(V) - \lambda_V\sigma(X,V)}{\lambda_V(1 - \lambda_V)} \tag{35}$$

$$= \frac{\sigma^2(X)\,\{\sigma^2(X)\sigma^2(V) - [\sigma(X,V)]^2\}}{\sigma(X,V)\,[\sigma^2(X) - \sigma(X,V)]} \,. \tag{36}$$

Equations 32 and 36 are the usual expressions for the Angoff error variances for V and X, respectively, for the case of V included in X (see Angoff, 1953; Petersen et al., 1989, p. 254).

Comments

Many of the results presented thus far have been provided implicitly or explicitly by others (e.g., Angoff, 1953; Feldt, 1975; and Woodruff, 1986). However, the particular form of some of the derivations presented here is somewhat novel and compact.

Also, strictly speaking, not all of the results that have been presented are required to derive the Levine-Angoff results considered subsequently. In particular, the reliabilities and Angoff error variances are not required *per se*, but they are useful in relating expressions of results to be presented with corresponding expressions provided by Angoff (1984), Kolen and Brennan (1987), and Petersen et al. (1989), among others.

Levine Observed-Score Method

The Levine observed-score method (elsewhere called the "equally reliable" method) for the common-item nonequivalent-populations design was originally developed by Levine (1955). Angoff (1984, p. 115) and Petersen et al. (1989, p. 254) also present descriptions of the method. Using a congeneric model, Woodruff (1986) derived a special case of the Levine-Angoff results. Subsequently, Kolen and Brennan (1987) derived a more general version of the Levine-Angoff results using a framework that explicitly incorporates the

synthetic group concept originally introduced by Braun and Holland (1982). The derivation outlined below integrates the Kolen and Brennan (1987) presentation and the classical congeneric model results presented previously.

Assume that a new test form X is administered to population 1 and an old test form Y is administered to population 2. (The adjectives "new" and "old" to describe forms X and Y, respectively, are used here for convenience only. There is nothing in the derivations that distinguishes between the "newness" or "oldness" of a form.) Also, assume that both populations take a common set of items, V, which may be distinct from X and Y or included in both X and Y. This is a description of the common-item non-equivalent populations design.

For this design, the two populations can be combined into a single population for defining the equating relationship. To address this issue Braun and Holland (1982) introduced the concept of a synthetic population. Statistics for populations 1 and 2 are proportionally weighted by w_1 and w_2, respectively, (i.e., $w_1 + w_2 = 1$ with $w_1, w_2 \geq 0$) to obtain statistics for the synthetic population.

For the Levine observed-score method, the linear equation for equating scores on X to the scale of Y is

$$\ell(X) = \frac{\sigma_s(Y)}{\sigma_s(X)} [X - \mu_s(X)] + \mu_s(Y) , \qquad (37)$$

where s indicates the synthetic population. For examinees in the synthetic population, the transformed observed scores on X [i.e., $\ell(X)$] have the same mean and standard deviation as the observed scores on Y.

Assumptions

Letting T_x, T_y, and T_v be true scores for X, Y, and V, respectively, Levine made the following three assumptions in deriving his results (see Kolen & Brennan, 1987, pp. 266-267):

(a) T_x and T_v correlate perfectly for both populations, and the same condition holds for T_y and T_v;

(b) the linear function of T_x on T_v is the same for both populations, and the same condition holds for T_y and T_v; and

(c) measurement error variance for X is the same for both populations, and the same condition holds for Y and V.

General Results

Letting subscripts designate populations, Kolen and Brennan (1987, see especially pp. 267, 268, and 272) show that under the Levine assumptions the four parameters in Equation 37 can be represented as:

$$\mu_s(X) = \mu_1(X) - w_2\gamma_1[\mu_1(V) - \mu_2(V)] , \quad (38)$$

$$\mu_s(Y) = \mu_2(Y) + w_1\gamma_2[\mu_1(V) - \mu_2(V)] , \quad (39)$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\gamma_1^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_1^2[\mu_1(V) - \mu_2(V)]^2 , \text{ and} \quad (40)$$

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\gamma_2^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_2^2[\mu_1(V) - \mu_2(V)]^2 ; \quad (41)$$

where the γ-terms are ratios of true-score standard deviations. In particular,

$$\gamma_1 = \sigma_1(T_x)/\sigma_1(T_v) \text{ and} \quad (42)$$

$$\gamma_2 = \sigma_2(T_y)/\sigma_2(T_v) . \quad (43)$$

(Angoff, 1987, and Brennan & Kolen, 1987, discuss and debate various issues with respect to choosing the weights w_1 and w_2.)

When the classical congeneric model is applied to obtain γ_1 and γ_2, the results discussed next are obtained for the external and internal anchor cases.

External Anchor

Substituting Equations 14 and 18 into Equation 42 we obtain

$$\gamma_1 = \sqrt{(\lambda_{x_1}/\lambda_{v_1})/(\lambda_{v_1}/\lambda_{x_1})}$$

$$= \lambda_{x_1}/\lambda_{v_1} , \tag{44}$$

where the subscript 1 is used to specify that the data are for examinees in population 1. In terms of variances and covariances, the effective test length parameters in Equation 44 are given by Equations 10 and 12. Therefore,

$$\gamma_1 = [\sigma_1^2(X) + \sigma_1(X,V)]/[\sigma_1^2(V) + \sigma_1(X,V)] . \tag{45}$$

Furthermore, the effective test length parameters in Equation 44 are also given by the slopes in Equations 11 and 13. Therefore,

$$\gamma_1 = \alpha_1(X|A)/\alpha_1(V|A) , \tag{46}$$

where $A = X + V$.

Corresponding equations for the old test form Y and population 2 can be obtained by substituting Y for X, 2 for 1, and $B = Y + V$ for $A = X + V$, in Equations 44-46 resulting in

$$\gamma_2 = \lambda_{y_2}/\lambda_{v_2} \tag{47}$$

$$= [\sigma_2^2(Y) + \sigma_2(Y,V)]/[\sigma_2^2(V) + \sigma_2(Y,V)] \tag{48}$$

$$= \alpha_2(Y|B)/\alpha_2(V|B) . \tag{49}$$

Equations 44 and 47 state that, for the Levine observed-score method with an external anchor, γ_1 and γ_2 (i.e., the ratio of the true-score standard deviations in Equations 42 and 43) are ratios of effective test lengths in populations 1 and 2, respectively.

Equations 45 and 48 are the most frequently reported expressions for the γ-terms (see, for example, Angoff, 1984, p. 115 and Kolen & Brennan, 1987, p. 272), but to this author these expressions lack the interpretability of the

effective test length ratios in Equations 44 and 47, and to some extent they lack the interpretability of the slope ratios in Equations 46 and 49.

Internal Anchor

Substituting Equations 29 and 33 into Equation 42, we obtain

$$\gamma_1 = \sqrt{\left[\frac{1}{\lambda_{v_1}(1 - \lambda_{v_1})}\right]\left[\frac{1 - \lambda_{v_1}}{\lambda_{v_1}}\right]}$$

$$= 1/\lambda_{v_1} . \qquad (50)$$

This, too, is a ratio of effective test lengths because, for the internal anchor case, the effective test length of X in population 1 is $\lambda_{x_1} = 1$ (see Equation 22), which is the numerator of Equation 50. Using Equation 28, an alternative expression for γ_1 is

$$\gamma_1 = 1/\alpha_1(V|X) , \qquad (51)$$

which is the expression provided by Angoff (1984, p. 115) and Kolen and Brennan (1987, p. 272). For the old form Y and population 2,

$$\gamma_2 = 1/\lambda_{v_2} \qquad (52)$$

$$= 1/\alpha_2(V|Y) . \qquad (53)$$

Comment

The derivation that has been outlined here of Levine's observed-score method integrates the Kolen and Brennan presentation of this method with the classical congeneric model results presented previously, with emphasis placed upon the interpretation of the γ-terms as ratios of effective test lengths. Certain aspects of the approach, results, and interpretations presented here are also provided by Angoff (1984, p. 114-115), Kolen and Brennan (1987), and Woodruff (1986). For example, Angoff's (1984) results are equivalent to those presented here when $w_1 = n_1/(n_1 + n_2)$ and $w_2 = n_2/(n_1 + n_2)$, where

n_1 and n_2 are sample sizes for populations 1 and 2, respectively. Woodruff's (1986) results are equivalent to those presented here when $w_1 = 1$.

Levine True-Score Method

Levine (1955) also developed another method for the common-item nonequivalent populations design. This second method is called Levine's true-score" method here. (Elsewhere, it is called Levine's "unequally reliable" method.) Angoff (1984, p. 115) and Petersen et al. (1989, p. 254) present descriptions of the method. The assumptions about true scores for this method are the same as those for the observed-score method. What distinguishes the methods is that the linear equation for the true-score method is expressed in terms of certain true-score quantities, rather than the observed-score quantities in Equation 37.

Specifically, for the true-score method, the basic linear equation is

$$g(T_x) = \frac{\sigma_s(T_y)}{\sigma_s(T_x)} [T_x - \mu_s(T_x)] + \mu_s(T_y) ,$$

where T_x designates the true score associated with a particular examinee's observed score. For examinees in the synthetic population, the transformed true scores on X [i.e., $g(T_x)$] have the same mean and standard deviation as the true scores on Y.

Clearly, however, examinees' true scores are never known. Therefore, the linear equation that is used in practice is

$$g(X) = \frac{\sigma_s(T_y)}{\sigma_s(T_x)} [X - \mu_s(T_x)] + \mu_s(T_y) \qquad (54)$$

$$= \frac{\sigma_s(T_y)}{\sigma_s(T_x)} [X - \mu_s(X)] + \mu_s(Y) , \qquad (55)$$

since true-score means equal observed-score means for the models considered here. Equation 54 or 55 is the Levine true-score equating function. The logic of using $g(X)$ rather than $g(T_x)$ is neither more nor less compelling than, for example, using observed scores in IRT true-score equating procedures (see Lord, 1980, p. 202). Note, in particular, that the transformed observed scores on X [i.e., $g(X)$] typically do not have the same standard deviation as the true scores on Y or the observed scores on Y.

General Results

Using the Kolen and Brennan (1987) approach, it can be shown that under Levine's assumptions:

$$\mu_s(T_x) = \mu_s(X) = \mu_1(X) - w_2\gamma_1[\mu_1(V) - \mu_2(V)] , \quad (56)$$

$$\mu_s(T_y) = \mu_s(Y) = \mu_2(Y) + w_1\gamma_2[\mu_1(V) - \mu_2(V)] , \quad (57)$$

$$\sigma_s^2(T_x) = \gamma_1^2\sigma_s^2(T_v) , \text{ and} \quad (58)$$

$$\sigma_s^2(T_y) = \gamma_2^2\sigma_s^2(T_v) . \quad (59)$$

where $\sigma_s^2(T_v) = w_1\sigma_1^2(V) + w_2\sigma_2^2(V) + w_1w_2[\mu_1(V) - \mu_2(V)]^2$.

Equations 56 and 57 are the same as the corresponding Equations 38 and 39, respectively, for the observed-score method. Equation 58 for the true-score variance of X in the synthetic population is derived in the Appendix, and Equation 59 can be derived in a similar manner.

Since the assumptions for both the observed-score and true-score methods are the same, in general the γ-terms in Equations 56-59 are the ratios of the true-score standard deviations in Equations 42 and 43. Furthermore, the γ-terms are the same as those derived previously using the classical congeneric models for the external and internal anchor cases. Thus, the γ's in Equations 56-59 are also interpretable as ratios of effective test lengths.

The simple form of Equations 58 and 59 leads to the slope of the linear equation for $g(X)$ in Equation 55 being

$$\sigma_s(T_y)/\sigma_s(T_x) = \gamma_2/\gamma_1 \qquad (60)$$

which is $(\lambda_{y_2}/\lambda_{v_2})/(\lambda_{x_1}/\lambda_{v_1})$ for the external anchor case, and $\lambda_{v_1}/\lambda_{v_2}$ for the internal anchor case. As shown below, the intercept for $g(X)$ can also be expressed relatively simply in terms of directly estimable parameters. Using Equations 56 and 57 with $v = \mu_1(V) - \mu_2(V)$, the intercept is

$$\begin{aligned}\mu_s(Y) - [\sigma_s(T_y)/\sigma_s(T_x)]\mu_s(X) &= \mu_2(Y) + w_1\gamma_2 v - (\gamma_2/\gamma_1)[\mu_1(X) - w_2\gamma_1 v] \\ &= \mu_2(Y) - (\gamma_2/\gamma_1)\mu_1(X) + \gamma_2(w_1 + w_2)v \ .\end{aligned}$$

Since $w_1 + w_2 = 1$, it follows that the intercept equals

$$[\mu_2(Y) - (\gamma_2/\gamma_1)\mu_1(X)] + \gamma_2[\mu_1(V) - \mu_2(V)] \ . \qquad (61)$$

Note that the slope and intercept do <u>not</u> depend on the weights, w_1 and w_2.

Replacing Equations 60 and 61 in Equation 55 we obtain

$$g(X) = (\gamma_2/\gamma_1)[X - \mu_1(X)] + \mu_2(Y) + \gamma_2[\mu_1(V) - \mu_2(V)] \ . \qquad (62)$$

Hence, $g(X)$ for Levine's true-score method is invariant with respect to weighting of populations 1 and 2 in forming the synthetic population, or we might say that the concept of a synthetic population is not necessary to conceptualize this method's results. Even so, it is sometimes useful to display Levine's true-score method in the form of Equations 55-59 to compare it with Levine's observed-score method in the form of Equations 37-41.

The usual presentation of results for the Levine true-score method is rather different from that presented here. Therefore, provided below are the "usual" Levine-Angoff results presented by Angoff (1984, p. 115), along with proofs of their equivalence to the results presented here (which assume, of course, the classical congeneric models discussed in this paper).

BEST COPY AVAILABLE

External Anchor

Angoff (1984, p. 115) states that, when V is an external anchor, the slope of g(X) is

$$\frac{\sigma_s(T_y)}{\sigma_s(T_x)} = \frac{\alpha_2(Y|V)\ \rho_1(V,V')}{\alpha_1(X|V)\ \rho_2(V,V')} \quad . \tag{63}$$

Using Equation 15 for $\rho_1(V,V')$ and the parallel equation for the reliability of V on population 2,

$$\frac{\sigma_s(T_y)}{\sigma_s(T_x)} = \frac{\alpha_2(Y|V)\ (\lambda_{V_1}/\lambda_{X_1})\sigma_1(X,V)/\sigma_1^2(V)}{\alpha_1(X|V)\ (\lambda_{V_2}/\lambda_{Y_2})\sigma_2(Y,V)/\sigma_2^2(V)} \quad .$$

Since $\alpha_1(X|V) = \sigma_1(X,V)/\sigma_1^2(V)$ and $\alpha_2(Y|V) = \sigma_2(Y,V)/\sigma_2^2(V)$, it follows that

$$\frac{\sigma_s(T_y)}{\sigma_s(T_x)} = \frac{\lambda_{Y_2}/\lambda_{V_2}}{\lambda_{X_1}/\lambda_{V_1}} \quad .$$

Finally, from Equations 44 and 47, we obtain the slope given by Equation 60.

Angoff (1984, p. 115) also states that the intercept of g(X) with V being an external anchor is

$$\mu_2(Y) - \frac{\sigma_s(T_y)}{\sigma_s(T_x)}\ \mu_1(X) + \frac{\alpha_2(Y|V)}{\rho_2(V,V')}\ [\mu_1(V) - \mu_2(V)] \quad . \tag{64}$$

Since $\alpha_2(Y|V) = \sigma_2(Y,V)/\sigma_2^2(V)$ and, by the parallel of Equation 15 for population 2, $\rho_2(V,V') = (\lambda_{V_2}/\lambda_{Y_2})\sigma_2(Y,V)/\sigma_2^2(V)$, it follows that $\alpha_2(Y|V)/\rho_2(V,V') = \lambda_{Y_2}/\lambda_{V_2} = \gamma_2$ by Equation 47. Therefore, since $\sigma_s(T_y)/\sigma_s(T_x) = \gamma_2/\gamma_1$, the intercept given by Equation 64 can be written as

$$\mu_2(Y) - (\gamma_2/\gamma_1)\mu_1(X) + \gamma_2[\mu_1(V) - \mu_2(V)] \quad ,$$

which is identical to the result in Equation 62.

Internal Anchor

For an internal anchor, Angoff (1984, p. 115) states that the slope of g(X) is

$$\sigma_s(T_y)/\sigma_s(T_x) = \alpha_1(V|X)/\alpha_2(V|Y) \quad . \tag{65}$$

For the development considered here, by Equations 58 and 59, 50 and 52, and 51 and 53, respectively, the slope is

$$\frac{\sigma_s(T_y)}{\sigma_s(T_x)} = \frac{\gamma_2}{\gamma_1} = \frac{\lambda_{v_1}}{\lambda_{v_2}} = \frac{\alpha_1(V|X)}{\alpha_2(V|Y)} ,$$

which equals Equation 65.

Angoff (1954, p. 115) also states that the intercept of g(X) for an internal anchor is

$$\mu_2(Y) - \frac{\sigma_s(T_y)}{\sigma_s(T_x)}\mu_1(X) + \frac{1}{\alpha_2(V|Y)}[\mu_1(V) - \mu_2(V)] . \qquad (66)$$

Using Equations 58, 59, and 53, we can rewrite Equation 66 as

$$\mu_2(Y) - (\gamma_2/\gamma_1)\mu_1(X) + \gamma_2[\mu_1(V) - \mu_2(V)] .$$

which is identical to the result in Equation 62.

First-Order Equity

For the Levine true-score method, a function relating true scores is applied to observed scores. As noted previously, the logic of doing so is somewhat less than compelling, and it is not clear how the converted scores on X, [i.e., g(X)] are comparable to scores on Y. Hanson (1990), however, has shown that Levine's true-score equating function (Equation 54) for the common-item nonequivalent-populations design results in first-order equity of the equated test scores under a particular parameterization of the classical congeneric model.

Before describing Hanson's modeling in more detail, we illustrate Hanson's approach for the much simpler case of the single group design and the Levine true-score equating function

$$g(X) = [\sigma(T_y)/\sigma(T_x)][X - \mu(T_x)] + \mu(T_y) . \qquad (67)$$

(With the single group design, no synthetic population is involved. Therefore, there are no subscripts on the parameters in Equation 67.)

Letting ψ be a function that relates true scores on X to true scores on Y, first-order, or weak, equity is defined as

$$\mathcal{E}[g(X)|\psi(T_x) = \tau] = \mathcal{E}(Y|T_y = \tau) \text{ for all } \tau . \quad (68)$$

Under this definition, the transformed score $g(X)$ is defined to be equivalent to Y if the expected value of the conditional distribution of $g(X)$ given $\psi(T_x) = \tau$ equals the expected value of the conditional distribution of Y given $T_y = \tau$. Divgi (1981), Morris (1982), and Yen (1983) consider first-order equity, which is a weaker case of the concept of equity first proposed by Lord (1980).

Consider the single group design with no common items, and assume that no context effects exist relative to the fact that examinees take both forms. For this design the congeneric model for test forms X and Y can be specified as

$$X = T_x + E_x = (\lambda_x T + \delta_x) + E_x \text{ and} \quad (69)$$

$$Y = T_y + E_y = (\lambda_y T + \delta_y) + E_y . \quad (70)$$

It follows that

$$T_y = (\lambda_y/\lambda_x)(T_x - \delta_x) + \delta_y = \psi(T_x) .$$

Consequently,

$$g(X) = [\delta_y - (\lambda_y/\lambda_x)\delta_x] + (\lambda_y/\lambda_x)X \quad (71)$$

satisfies the condition of first-order equity in Equation 68, because the expected value of $g(X)$ in Equation 71, given τ, equals the expected value of Y given τ, for all τ.

To show that $g(X)$ in Equation 67 satisfies first-order equity, it is sufficient to show that it equals Equation 71. From Equations 69 and 70,

$$\mu(T_x) = \lambda_x\mu(T) + \delta_x , \quad \mu(T_y) = \lambda_y\mu(T) + \delta_y ,$$

$$\sigma(T_x) = \lambda_x\sigma(T) , \text{ and } \sigma(T_y) = \lambda_y\sigma(T) .$$

Replacing these equations in Equation 67 gives

$$g(X) = \frac{\lambda_y}{\lambda_x}\left[X - \lambda_x\mu(T) - \delta_x\right] + \lambda_y\mu(T) + \delta_y$$

$$= \frac{\lambda_y}{\lambda_x}X + \left[\lambda_y\mu(T) + \delta_y - \lambda_y\mu(T) - \frac{\lambda_y}{\lambda_x}\delta_x\right]$$

$$= \left[\delta_y - \frac{\lambda_y}{\lambda_x}\delta_x\right] + \frac{\lambda_y}{\lambda_x}X ,$$

which is identical to Equation 71.

The same type of logic has been applied by Hanson (1990) in the much more complicated context of the common-items nonequivalent-populations design. Specifically, except for slight notational differences, Hanson (1990) uses the following congeneric model for the test forms and common items:

$$H_2 = T_2 + E_{h_2}, \tag{72}$$

$$V_2 = (\lambda_v T_2 + \delta_v) + E_{v_2}, \tag{73}$$

$$K_1 = (\lambda_k T_1 + \delta_k) + E_{k_1}, \quad \text{and} \tag{74}$$

$$V_1 = (\lambda_v T_1 + \delta_v) + E_{v_1}, \tag{75}$$

where 1 and 2 designate populations, T_i $(i = 1,2)$ is the true-score random variable corresponding to the observable score H_i, $Y = H + V$, and $X = K + V$. Further, the error variances are assumed to satisfy the assumptions of the classical congeneric model, and the constraints imposed are $\lambda_h = 1$ and $\delta_h = 0$. Given this modeling, Hanson (1990) shows that Levine's true-score equating procedure satisfies first-order equity, for both internal and external sets of common items.

Note that Hanson's modeling of congeneric forms in Equations 72-75 differs considerably from that discussed in previous sections of this paper. In particular, Equations 72-75 directly relate V to both X and Y in a single

model. In other words, Equations 72-75 constitute one model with one λ_V term for V, whereas in previous sections the common-items nonequivalent-populations design was framed in terms of separate congeneric models for the two forms, which involves two effective test length parameters for V.

Summary and Discussion

In this paper, two different parameterizations of a classical congeneric model have been introduced to model explicitly the variables in the Levine observed-score and true-score linear equating procedures, for the external and internal anchor cases. The models differ in the constraints imposed on the effective test length parameters, λ_X and λ_V, as well as assumptions made about one covariance term, $\sigma(E_X, E_V)$. With an external anchor the model employs the constraint $\lambda_X + \lambda_V = 1$ and assumes $\sigma(E_X, E_V) = 0$, whereas with an internal anchor λ_X is set to 1, and it is assumed that $\sigma(E_X, E_V) = \lambda_V \sigma^2(E)$. Using these two parameterizations, relatively simple expressions are easily obtained for true-score variances, reliabilities, and error variances. Further, the error variances are equal to the so-called "Angoff error variances."

Applying these two parameterizations of the classical congeneric model with the Levine assumptions leads to general equations (for both of Levine's procedures and both the external and internal anchor cases) that involve ratios of effective test length parameters. This aids interpretation.

The derived results are summarized in Table 1, where $\ell(X)$ and $g(X)$ are the linear functions for the observed-score and true-score methods, respectively. There are similarities between the expression of some of the results in Table 1 and other expressions of results for the Levine procedures (notably, Angoff, 1984, p. 115, and Kolen & Brennan, 1987, p. 272). For

example, as in Kolen and Brennan (1987), results are expressed in terms of synthetic population weights, means and variances that are directly observable, and certain γ-terms. (Kolen & Brennan, however, provide results for the observed-score case, only.) Also, for $w_1 = n_1/(n_1 + n_2)$ and $w_2 = n_2/(n_1 + n_2)$ the results in Table 1 are algebraically equivalent to those presented by Angoff (1984).

There are, however, several differences between the expression of results summarized in Table 1 and other expressions. First, the γ-terms are all expressed as ratios of effective test length parameters for the two parameterizations of the classical congeneric model used in this paper. This fact enhances the interpretability of the γ-terms. For example, with an exclusive anchor, it is evident that γ_1 increases as the effective test length of X increases relative to V in population 1. Second, the γ-terms are the same for both the observed-score and true-score methods. Third, the effective test length parameters are all slopes in a particular linear regression. In general $\lambda_{F_i} = \alpha_i(F|*)$ where i = 1 or 2, F is X, Y, or V, and * is a total score involving F. Fourth, the linear function for the true-score method, g(X), can be obtained using expressions for synthetic group means and variances that involve synthetic population weights, but g(X) itself is blind to such weights. This is a notable difference between the observed-score and true-score methods--a difference that has not been reported previously.

The assumptions about true scores and error variances for both of the Levine methods are the same. What distinguishes the methods is the nature of the linear functions. For the observed-score method, the linear function relates converted observed scores on X to scores on Y. For the true-score method, however, the basic linear function relates true scores, but it is applied to observed scores. Consequently, for the true-score method, it is

not clear how the converted scores on X are in any sense comparable to scores on Y. Recently, however, Hanson (1990) has shown that Levine's true-score method satisfies the condition of first-order equity under a particular parameterization of the classical congeneric model. Of course, this does not necessarily mean that Levine's true-score method is preferable to Levine's observed-score method, but Hanson's proof casts new light on Levine's true-score method.

Although the two Levine methods are not properly distinguished in terms of being derived under assumptions about equally reliable and unequally reliable tests, there is a relationship between the two methods that involves reliabilities. In particular, if there exists a particular synthetic population in which X and Y are equally reliable [i.e., $\rho_s(X,X') = \rho_s(Y,Y')$ for a particular w_1 (and $w_2 = 1 - w_1$)], then

$$\frac{\sigma_s(T_y)}{\sigma_s(T_x)} = \frac{\sigma_s(Y)}{\sigma_s(X)} \frac{\rho_s(Y,Y')}{\rho_s(X,X')} = \frac{\sigma_s(Y)}{\sigma_s(X)} ,$$

$\ell(X) = g(X)$ for this synthetic population, and for both methods the converted scores on X for the synthetic population will have the same mean and variance as the scores on Y. Note that this equivalence does <u>not</u> necessarily hold for every synthetic population, however.

Sometimes the following question is asked: "When tests are equally reliable, why doesn't Levine's unequally reliable procedure give the same results as Levine's equally reliable procedure?" This seemingly sensible question, however, is somewhat misleading and ambiguous. It is misleading because, as shown in this report, the generalized version of the Levine-Angoff results summarized in Table 1 can be derived without any assumptions about reliability. The so-called "equally reliable" procedure is simply the observed-score method denoted $\ell(X)$, and the so-called "unequally reliable"

procedure is simply the true-score method denoted $g(X)$. The question is ambiguous because it fails to recognize the role of the synthetic population in obtaining $\ell(X)$. For example, suppose $\rho_1(X,X') = \rho_2(Y,Y')$, which implies that X and Y are equally reliable for the populations that actually took X and Y. It does not follow, however, that $\rho_s(X,X') = \rho_s(Y,Y')$ for the particular synthetic population actually used. Thus, it is quite possible for forms to be equally reliable in some sense without having $\ell(X) = g(X)$.

Levine's (1955) methods make assumptions about true scores. Consequently, to apply these methods, one must employ some model that relates observed and true scores. Levine employed classical test theory assumptions, and expressed many of his results in terms of reliability coefficients. However, he gave only limited consideration to how such coefficients might be estimated. Angoff's (1984) results are based on estimating these coefficients using observed variances and Angoff's (1953) error variances. In this paper, two specific classical congeneric models are used to derive results for Levine's methods. These results can be viewed as more general versions of Angoff's results, although they are derived and expressed differently.

Since Levine's methods require some model for the relationship between observed and true scores, models other than the classical congeneric model could lead to different results. In particular, the multi-factor congeneric model discussed by Feldt and Brennan (1989, p. 111), or one or more models in generalizability theory, might be employed with Levine's methods. The principal point is that improved estimates of true-score variances, error variances, or reliabilities could lead to improved results. Also, improvements might result from relaxing one or more of Levine's assumptions.

References

Angoff, W. H. (1953). Test reliability and effective test length. Psychometrika, 18, 1-14.

Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service. (Reprint of chapter in R. L. Thorndike (Ed.), Educational measurement (2nd Ed.). Washington, DC: American Council on Education, 1971.)

Angoff, W. H. (1987). Technical and practical issues in equating: A discussion of four papers. Applied Psychological Measurement, 11, 291-300.

Braun, H. I., & Holland, P. W. (1982). Observed score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), Test equating. New York: Academic Press.

Brennan, R. L., & Kolen, M. J. (1987). A reply to Angoff. Applied Psychological Measurement, 11, 301-306.

Divgi, D. R. (1981). Model-free evaluation of equating and scaling. Applied Psychological Measurement, 5, 203-208.

Feldt, L. S. (1975). Estimation of the reliability of a test divided into two parts of unequal length. Psychometrika, 40, 557-561.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), Educational measurement (3rd ed.), 105-146. New York: Macmillan.

Hanson, B. A. (in press). A note on Levine's formula for equating unequally reliable tests using data from the common item nonequivalent groups design. Journal of Educational Statistics.

Kolen, M. J., & Brennan, R. L. (1987). Linear equating models for the common-item nonequivalent-populations design. Applied Psychological Measurement, 11, 263-277.

Levine, R. (1955). Equating the score scales of alternative forms administered to samples of different ability (RB-55-23). Princeton, NJ: Educational Testing Service.

Lord, F. M. (1980). Applications of item response theory to practical measurement problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacCann, R. G. (1990). Derivations of observed score equating methods that cater to populations differing in ability. Journal of Educational Statistics, 15, 146-170.

Morris, C. N. (1982). On the foundations of test equating. In P. W. Holland & D. B. Rubin (Eds.), Test equating, 169-197. New York: Wiley.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), Educational measurement (3rd Ed.), 221-262. New York: American Council on Education, and Macmillan.

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. Review of Educational Research, 4, 663-705.

Woodruff, D. J. (1986). Derivations of observed score linear equating methods based on test score models for the common item nonequivalent populations design. Journal of Educational Statistics, 11, 245-357.

Woodruff, D. J. (1989). A comparison of three linear equating methods for the common-item nonequivalent populations design. Applied Psychological Measurement 13, 257-261.

Yen, W. (1983). Tau-equivalent and equipercentile equating. Psychometrika, 48, 353-369.

Appendix

Proof that $\sigma_s^2(T_X) = \gamma_1^2\sigma_s^2(T_V)$

In general, it is easy to show that

$$\sigma_s^2(T_X) = w_1\sigma_1^2(T_X) + w_2\sigma_2^2(T_X) + w_1w_2[\mu_1(T_X) - \mu_2(T_X)]^2 . \quad (A1)$$

For the classical congeneric model $\mu_1(T_X) = \mu_1(X)$ and $\mu_2(T_X) = \mu_2(X)$. Also, under Levine's assumptions Kolen and Brennan (1987, Equation 32) show that

$$\mu_2(X) = \mu_1(X) - [\sigma_1(T_X)/\sigma_1(T_V)][\mu_1(V) - \mu_2(V)] .$$

It follows that

$$\sigma_s^2(T_X) = w_1\sigma_1^2(T_X) + w_2\sigma_2^2(T_X) + w_1w_2[\sigma_1^2(T_X)/\sigma_1^2(T_V)][\mu_1(V) - \mu_2(V)]^2$$

$$= \frac{\sigma_1^2(T_X)}{\sigma_1^2(T_V)} \left\{ w_1\sigma_1^2(T_V) + w_2\frac{\sigma_1^2(T_V)}{\sigma_1^2(T_X)}\sigma_2^2(T_X) + w_1w_2[\mu_1(V) - \mu_2(V)]^2 \right\} \quad (A2)$$

Under the Levine assumptions, the slope of the linear function of T_X on T_V is the same in populations 1 and 2. This means that

$$\sigma_1(T_X)/\sigma_1(T_V) = \sigma_2(T_X)/\sigma_2(T_V) . \quad (A3)$$

Applying Equation A3 to the second term in braces in Equation A2 gives

$$\sigma_s^2(T_X) = \frac{\sigma_1^2(T_X)}{\sigma_1^2(T_V)} \left\{ w_1\sigma_1^2(T_V) + w_2\sigma_2^2(T_V) + w_1w_2[\mu_1(V) - \mu_2(V)]^2 \right\} .$$

The term in braces is $\sigma_s^2(T_V)$, and by Equation 42 $\sigma_1^2(T_X)/\sigma_1^2(T_V) = \gamma_1^2$. Thus

$$\sigma_s^2(T_X) = \gamma_1^2\sigma_s^2(T_V) ,$$

as was to be proved.

Author Note

This is a revised version of a paper presented at the Annual Meeting of the Psychometric Society in Princeton, NJ, in June 1990. The author appreciates many helpful comments and suggestions made by Bradley A. Hanson, Michael T. Kane, Michael J. Kolen, and David J. Woodruff.

Table 1

Equations for Levine's Observed-Score [$\ell(X)$] and True-Score [$g(X)$] Methods

$$\ell(X) = [\sigma_s(Y)/\sigma_s(X)][X - \mu_s(X)] + \mu_s(Y)$$

$$g(X) = [\sigma_s(T_y)/\sigma_s(T_x)][X - \mu_s(X)] + \mu_s(Y)$$

$$= (\gamma_2/\gamma_1)[X - \mu_1(X)] + \mu_2(Y) + \gamma_2[\mu_1(V) - \mu_2(V)]$$

$$\mu_s(X) = \mu_1(X) - w_2\gamma_1[\mu_1(V) - \mu_2(V)]$$

$$\mu_s(Y) = \mu_2(Y) + w_1\gamma_2[\mu_1(V) - \mu_2(V)]$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\gamma_1^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_1^2[\mu_1(V) - \mu_2(V)]^2$$

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\gamma_2^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_2^2[\mu_1(V) - \mu_2(V)]^2$$

$$\sigma_s^2(T_x) = \gamma_1^2\sigma_s^2(T_v)$$

$$\sigma_s^2(T_y) = \gamma_2^2\sigma_s^2(T_v) .$$

where $\sigma_s^2(T_v) = w_1\sigma_1^2(T_v) + w_2\sigma_2^2(T_v) + w_1w_2[\mu_1(V) - \mu_2(V)]^2$

External Anchor (A = X + V, B = Y + V) (Classical congeneric model)

$$\gamma_1 = \frac{\lambda_{x_1}}{\lambda_{v_1}} = \frac{\alpha_1(X|A)}{\alpha_1(V|A)} = \frac{\sigma_1^2(X) + \sigma_1(X,V)}{\sigma_1^2(V) + \sigma_1(X,V)}$$

$$\gamma_2 = \frac{\lambda_{y_2}}{\lambda_{v_2}} = \frac{\alpha_2(Y|B)}{\alpha_2(V|B)} = \frac{\sigma_2^2(Y) + \sigma_2(Y,V)}{\sigma_2^2(V) + \sigma_2(Y,V)}$$

Internal Anchor (Classical congeneric model)

$$\gamma_1 = 1/\lambda_{v_1} = 1/\alpha_1(V|X) = \sigma_1^2(X)/\sigma_1(X,V)$$

$$\gamma_2 = 1/\lambda_{v_2} = 1/\alpha_2(V|Y) = \sigma_2^2(Y)/\sigma_2(Y,V)$$

Note. For Tucker's method use $\gamma_1 = \alpha_1(X|V)$ and $\gamma_2 = \alpha_2(Y|V)$ in $\ell(X)$ for both the internal and external anchor cases.